HEALTHCARE COST AND UTILIZATION PROJECT — HCUP
A FEDERAL-STATE-INDUSTRY PARTNERSHIP IN HEALTH DATA
Sponsored by the Agency for Healthcare Research and Quality

General Information for

The HCUP Nationwide Inpatient Sample (NIS), Release 6, 1997

Agency for Healthcare Research and Quality
2101 East Jefferson Street, Suite 605
Rockville, Maryland 20852

phone: (301) 594-3075
fax: (301) 594-2166
Internet: hcup@ahrq.gov

NIS Data and Documentation Distributed by:
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161

phone: 800-553-6847
(703) 605-6000

Table of Contents

Table of Tables

Figures

HCUP NATIONWIDE INPATIENT SAMPLE (NIS)
SUMMARY OF DATA USE LIMITATIONS

******* REMINDER *******

Authorized users of HCUP data have agreed to the following limitations:†

- Cannot use the data without signing a data use agreement‡
- Cannot use the data for any purpose other than research or aggregate statistical reporting
- Cannot re-release any data to unauthorized users
- Cannot identify nor attempt to identify any individual
- Cannot link the data to data from another source that identifies individuals
- Cannot report information that could identify individual establishments
- Cannot use the data concerning individual establishments for commercial or competitive purposes involving those establishments
- Cannot use the data to determine rights, benefits, or privileges of individual establishments
- Cannot identify nor attempt to identify any establishment when its identity has been concealed on the database
- Cannot contact establishments included in the data
- Cannot attribute to data contributors any conclusions drawn from the data
- Must acknowledge the "Healthcare Cost and Utilization Project, 1988-1997 (HCUP)" in reports

Any violation of the limitations in the data use agreement is punishable under Federal law by a fine of up to $10,000 and up to 5 years in prison. Violations may also be subject to penalties under State statutes.

---

† Specific provisions are detailed in the Data Use Agreement for HCUP Nationwide Inpatient Sample for Release.
‡ A copy of this agreement is provided in Technical Supplement 1: Data Use Agreement for HCUP Nationwide Inpatient Sample for Release (See NIS Documentation on CD #6).

USER FEEDBACK AND TECHNICAL SUPPORT

We would like to receive your feedback on the HCUP NIS data products.

Our Internet address for user feedback and technical support is:

hcup@ahrq.gov

You may also send comments or concerns in writing to:

Agency for Healthcare Research and Quality
Healthcare Cost and Utilization Project (HCUP)
2101 E. Jefferson Street, Suite 605
Rockville, MD 20852

phone: (301) 594-3075
fax: (301) 594-2166

The Agency for Healthcare Research and Quality and the staff of the Healthcare Cost and Utilization Project (HCUP) thank you for purchasing the HCUP Nationwide Inpatient Sample (NIS) Release 6 on CD-ROM.

HCUP Nationwide Inpatient Sample (NIS) Release 6, 1997

ABSTRACT

The Nationwide Inpatient Sample (NIS) Release 6 is part of the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality, formerly the Agency for Health Care Policy and Research.

NIS Release 6 approximates a 20-percent sample of U.S. community hospitals for 1997. It is based on a stratified probability sample of hospitals, with sampling probabilities proportional to the number of U.S. community hospitals in each stratum.

NIS Release 6 is drawn from 22 States and contains information on all inpatient stays from over 1,000 hospitals, totaling about 7.1 million records in 1997. Inpatient stay records include clinical and resource use information typically available from discharge abstracts. Hospital and discharge weights are provided for producing national and State-level estimates. The NIS can be linked to hospital-level data from the American Hospital Association's Annual Survey of Hospitals and county-level data from the Bureau of Health Professions' Area Resource File (except for hospitals in Georgia, Hawaii, Kansas, South Carolina, and Tennessee).

Access to the NIS is open to users who sign data use agreements. Uses are limited to research and aggregate statistical reporting.

INTRODUCTION TO THE HCUP NATIONWIDE INPATIENT SAMPLE (NIS)

There are five sections to the Introduction to the HCUP Nationwide Inpatient Sample (NIS), which are:

- OVERVIEW OF NIS DATA
- CONTENTS OF CD-ROM SET CONTAINING NIS RELEASE 6
- NIS DATA ELEMENTS
- GETTING STARTED
- OTHER HCUP PRODUCTS

OVERVIEW OF NIS DATA

The HCUP Nationwide Inpatient Sample contains all-payer data on hospital inpatient stays from selected States. Each year of the NIS provides information on approximately 5 million to 7.1 million inpatient stays from over 900 hospitals. All discharges from sampled hospitals are included in the NIS database.

The NIS contains patient-level clinical and resource use information included in a typical discharge abstract. The NIS can be linked directly to hospital-level data from the American Hospital Association (AHA) Annual Survey of Hospitals and to county-level data from the Health Resources and Services Administration Bureau of Health Professions' Area Resource File (ARF), except for hospitals from Georgia, Hawaii, Kansas, South Carolina, and Tennessee.

The NIS is designed to approximate a 20-percent sample of U.S. community hospitals, defined by the AHA to be "all nonfederal, short-term, general, and other specialty hospitals, excluding hospital units of institutions." Included among community hospitals are specialty hospitals such as obstetrics-gynecology, ear-nose-throat, short-term rehabilitation, orthopedic, and pediatric. Excluded are long-term hospitals, psychiatric hospitals, and alcoholism/chemical dependency treatment facilities.

Six releases of NIS data are currently available, as shown in Table 1. Each release of the NIS includes:

- Data in ASCII format on CD-ROM
- Patient-level hospital discharge abstract data for 100 percent of discharges from a sample of hospitals in participating States
- 5 million to 7.1 million inpatient records per year
- 800-1,000 hospitals per year
- Two 10% Subsamples of discharges from all NIS hospitals
- Hospital Weights File to produce national estimates and to link the NIS to data from the American Hospital Association Annual Survey of Hospitals
- Hospital-Level Casemix Counts – small data files useful for generating estimates of discharges, length of stay, and charges, by diagnosis-related group or diagnosis and procedure categories
- NIS Documentation and NIS Tools, also on CD-ROM – including programming source code for loading ASCII data into SAS and SPSS, and value labels

Table 1. Summary of NIS Releases 1-6

	Data from	Media/format options	Structure of Releases
Release 1	• 1988-1992 • 8 States in 1988 • 11 States in 1989-1992	On CD-ROM, in ASCII format	5 years of NIS data in a 26-CD set 4 CDs per year (one quarter per disk) Two 10% Subsamples of discharges for each year
		On magnetic tape, in EBCDIC or SAS transport format	Each year sold separately
Release 2	• 1993 • 17 States	On CD-ROM, in ASCII format	1 year of data in a 6-CD set, structured for use on microcomputers Two 10% Subsamples of discharges
Release 3	• 1994 • 17 States		
Release 4	• 1995 • 19 States		
Release 5	• 1996 • 19 States		
Release 6	• 1997 • 22 States		

NIS Data Sources, Hospitals, and Inpatient Stays

Table 2 summarizes the data sources, number of hospitals, and number of inpatient stays in NIS data.

Table 2. Summary of NIS Data Sources, Hospitals and Inpatient Stays for Releases 1-6

Calendar year	Data sources	Number of hospitals	Number of inpatient stays
NIS Release 1			
1988	CA CO FL IL IA MA NJ WA	759	5,265,756
1989	AZ CA CO FL IL IA MA NJ PA WA WI *(added AZ, PA, WI)*	882	6,110,064
1990	AZ CA CO FL IL IA MA NJ PA WA WI *(no change)*	871	6,268,515
1991	AZ CA CO FL IL IA MA NJ PA WA WI *(no change)*	859	6,156,188
1992	AZ CA CO FL IL IA MA NJ PA WA WI *(no change)*	856	6,195,744
NIS Release 2			
1993	AZ CA CO CT FL IL IA KS MD MA NJ NY OR PA SC WA WI *(added CT, KS, MD, NY, OR, SC)*	913	6,538,976
NIS Release 3			
1994	AZ CA CO CT FL IL IA KS MD MA NJ NY OR PA SC WA WI *(no change)*	904	6,385,011
NIS Release 4			
1995	AZ CA CO CT FL IL IA KS MD MA MO NJ NY OR PA SC TN WA WI *(added MO, TN)*	938	6,714,935
NIS Release 5			
1996	AZ CA CO CT FL IL IA KS MD MA MO NJ NY OR PA SC TN WA WI *(no change)*	906	6,542,069
NIS Release 6			
1997	AZ CA CO CT FL GA HI IL IA KS MD MA MO NJ NY OR PA SC TN UT WA WI *(added GA, HI, UT)*	1,012	7,148,420

CONTENTS OF CDROM SET CONTAINING NIS RELEASE 6

There are two types of files included in each NIS release: 1) data files and 2) documentation and tools files. The following is a brief description of these types of files:

1) Data Files - five types of ASCII formatted data files are included:

   Inpatient Core Files: These inpatient discharge-level files contain data for 100% of the discharges from a sample of hospitals in participating states. Core File A contains the most commonly used data elements and Core Files B and C contain less commonly used data elements.

   Normalized Inpatient Diagnoses File: This diagnosis-level file contains one record per diagnosis per discharge record, along with the data element necessary to link to the discharge record in the Core Files; each discharge record can have up to 15 diagnoses linked to it.

   Normalized Inpatient Procedures File: This procedure-level file contains one record per procedure per discharge record, along with the data element necessary to link to the discharge record in the Core Files; each discharge record can have up to 15 procedures linked to it.

   Subsample Inpatient Core Files: Each of these discharge-level files contain all data elements from the Core Files, for a 10% subsample of the NIS; these can be combined to create a 20% NIS subsample.

   Weights File: This hospital-level file contains weights and variance estimation data elements, as well as hospital name, location and linkage data elements.

2) Documentation and Tools Files

   Documentation: Complete file documentation, including codebooks and variable notes, is provided in a series of Portable Document Format (.PDF) files. See Section on Getting Started.

   SAS source code: Code is included for the format library for the variables, for loading ASCII data into SAS format, and for reassembling the diagnosis and procedure normalized files into original format.

   SPSS source code: Code is included for the variable library, for loading ASCII data into SPSS format, and for reassembling the diagnosis and procedure normalized files into original format.

   Labels: Labels for the Clinical Classifications Software (CCS), formerly called the Clinical Classifications for Health Policy Research (CCHPR), and for the Diagnosis-Related Groups (multiple versions).

   Record Layouts: Record layouts for all data files.

   Hospital-level casemix and counts files: Contains Hospital-Level Casemix Counts files with total number of discharges, sum of edited LOS, sum of edited TOTCHG, number of discharges reporting LOS, number of discharges reporting TOTCHG.

The data files are provided on a set of five CD-ROMs and the documentation and tools are on CD # 6; The contents of each CD are shown in Table 3.

Because of the file size this year, the data on three of the CD-ROMs had to be compressed into self-extracting zipped files. These CDs are noted below and are characterized by having the data in files named ***NIS97CDn.exe*** where the n denotes the CD number. To decompress these files, you should follow these steps:

1. Create a directory for the NIS.
2. Copy the self-extracting file from the CD into the new directory.
3. Unzip the file, by typing its name (e.g., NIS97CD1, NIS97CD3, or NIS97CD4) within DOS or clicking on the name within Windows Explorer.

The ASCII files mentioned below will then be uncompressed into this directory. After the file is uncompressed, the .exe file can be deleted.

Table 3. Contents of NIS, Release 6, 1997, on CD-ROM – 6-CD Set

CD #1. (NIS97CD1.exe)

1997, Inpatient Stay Core File A, includes most commonly used data elements†:

- Patient age, sex, race
- Expected primary source of payment
- Discharge status
- Discharge quarter
- Length of stay (edited)
- Total charges (edited)
- Total number of diagnoses and procedures
- Discharge weight
- Linkages to other core files and to Hospital Weights File
- Principal diagnosis and procedure (with validity indicators)
- Classifications based on diagnoses and procedures (CCS, DRG, MDC)

1997, Hospital Weights File, includes:

- Hospital name, location, and linkages to core files, AHA Survey, and Area Resource File
- Weights and other variables required to calculate weighted estimates and variances

CD #2.

1997, Inpatient Stay Core File B, includes less commonly used data elements:

- Admission source
- Admission type
- Neonatal/maternal indicator
- Physician identifiers (encrypted)
- Length of stay (unedited)
- Total charges (unedited)
- Day of principal procedure
- Linkages to other core files

CD #3. (NIS97CD3.exe)

1997, Inpatient Stay Core File C, includes less commonly used data elements:

- Admission day of week
- Admission month
- Data source identifiers (State and hospital)
- Expected sources of payment
- Linkages to other core files and Area Resource File

1997, Inpatient Stay Core File PR, normalized procedure file, includes:

- All procedures, validity indicators, and linkages to other core files

CD #4. (NIS97CD4.exe)

1997, Inpatient Stay Core File DX, normalized diagnosis file, includes:

- All diagnoses, validity indicators, and linkages to other core files

CD #5.

Two 10% Subsamples

- All data elements, including discharge weight, adjusted for subsampling
- These two non-overlapping samples can be combined to form one 20% sample of NIS discharges

CD #6.

1997, Hospital Weights File (same as on CD #1)
1997, Hospital-Level Casemix Counts by DRG10, DCCS1, PCCS1
NIS Documentation and Tools

† See Table 4 for explanation of data elements.

NIS Documentation and NIS Tools provide important resources for the user. Refer to these resources to understand the structure and content of the NIS and to aid in using the NIS.

NIS Documentation
- on CD-ROM #6
- in *.pdf (*portable document format*) files. Files with the *.pdf extension can be viewed, searched, and printed using the Adobe™ Acrobat™ Reader (see "Getting Started")
- in printed copy for magnetic tape users (Release 1 only)

NIS Tools, developed to assist users, are available in ASCII text files:
- on CD-ROM #6
- on diskette for magnetic tape users (Release 1 only)
- SAS and SPSS tools:
  - source code to load EBCDIC/ASCII data into SAS and SPSS formats
  - SAS and SPSS source code to create value labels
  - source code to reassemble diagnosis (DX) and procedure (PR) normalized files into original format
- Labels for Clinical Classifications Software (formerly Clinical Classifications for Health Policy Research, Version 2)
- Labels for Diagnosis-Related Groups, multiple versions

See "Getting Started," later in this document, for more information on documentation and tools.

NIS DATA ELEMENTS

All releases of the NIS contain two types of data: inpatient stay records and hospital information with weights. Later releases of the NIS (from Release 2 on) use the *Quick Reference* format, which may reduce the need to use multiple CDs for some projects. The Quick Reference format is described in detail under *General Information* (GENINFO.pdf).

The following table illustrates the file structure of the NIS for Releases 2 through 6 and identifies the data elements that can be found in each file.

Table 4. File Locations of Data Elements in the HCUP NIS

Data element	Description (numbers in brackets indicate variable coding)
CD #1	**Compressed file: NIS97CD1.EXE contains:**
Inpatient Stay Core File A (filename: NIS97_A.ASC)	
HOSPID	HCUP hospital number (links to Hospital Weights file)
SEQ	Unique sequence number indicates the order in which the data file is sorted (links across files)
AGE	Age in years at admission
AGEDAY	Age in days (coded only when the age in years is less than 1)
RACE	Race includes (1) white, (2) black, (3) Hispanic, (4) Asian or Pacific Islander, (5) Native American, (6) other
SEX	Sex includes (1) male, (2) female
ZIPINC4	Median income for patient's ZIP Code: (1) $0-25,000, (2) $25,001-$30,000, (3) $30,001-$35,000, (4) $35,001+
DX1	Principal diagnosis
DXV1	Validity flag for principal diagnosis – indicates (0) valid, (1) invalid codes
NDX	Number of diagnoses coded on the original record
PR1	Principal procedures
PRV1	Validity flag for principal procedure – indicates (0) valid, (1) invalid codes
NPR	Number of procedures coded on the original record
DRG	Diagnosis Related Group (DRG) in use on discharge date
MDC	Major Diagnosis Category (MDC) in use on discharge date
DRGVER	Grouper version in use on discharge date
DRG10	DRG Version 10 (October 1992)

Table 4. File Locations of Data Elements in the HCUP NIS (Continued)

Data element	Description (numbers in brackets indicate variable coding)
MDC	Major Diagnosis Category (MDC) in use on discharge date
DRGVER	Grouper version in use on discharge date
DRG10	DRG Version 10 (October 1992)
MDC10	MDC Version 10 (October 1992)
DCCS1	Clinical Classifications Software (CCS) for principal diagnosis
PCCS1	CCS for principal procedure
DQTR	Discharge quarter
LOS	Length of stay, edited
DIED	Indicates in-hospital death: (0) did not die during hospitalization, (1) died during hospitalization
DISP	Disposition of patient (discharge status): (1) routine, (2) short term hospital, (3) skilled nursing facility, (4) intermediate care, (5) another type of facility, (6) home health care, (7) against medical advice, (20) died
PAY1	Expected primary payer, uniform: (1) Medicare, (2) Medicaid, (3) private including HMO, (4) self-pay, (5) no charge, (6) other
TOTCHG	Total charges, edited
DISCWT_U	Weight to discharges in universe

Hospital Weights File (filename: NIS97WT.ASC)-- see CD #6 for contents

CD #2

Inpatient Stay Core File B (filename: NIS97_B.ASC)

SEQ	Unique sequence number indicates the order in which the data file is sorted (links across files)
MDID_S	Synthetic attending physician number
SURGID_S	Synthetic primary surgeon number
DSHOSPID	Hospital number as received from the data source
ZIPINC8	Median income for patient's ZIP Code: (1) $0-$15,000, (2) $15,001-$20,000, (3) $20,001-$25,000, (4) $25,001-$30,000, (5) $30,001-$35,000, (6) $35,001-$40,000, (7) $40,001-$45,000, (8) $45,001+
NEOMAT	Neonatal/maternal flag: (0) not maternal or neonatal, (1) maternal diagnosis or procedure, (2) neonatal diagnosis, (3) maternal and neonatal on same record
LOS_X	Length of stay, unedited
PRDAY1	Days from admission of principal procedure

Table 4. File Locations of Data Elements in the HCUP NIS (Continued)

Data element	Description (numbers in brackets indicate variable coding)
ASOURCE	Admission source: (1) ER, (2) another hospital, (3) another facility incl long term care, (4) court/law enforcement, (5) routine/birth/other
ATYPE	Admission type: (1) emergency, (2) urgent, (3) elective, (4) newborn, (5) delivery, (6) other
TOTCHG_X	Total charges, unedited
CD #3	**Compressed file: NIS97CD3.EXE contains**:
Inpatient Stay Core File C (filename: NIS97_C.ASC)	
SEQ	Unique sequence number indicates the order in which the data file is sorted (links across files)
PROCESS	Processing number assigned for tracking records throughout data processing
SEQ_SID	Unique sequence number assigned for processing original data file (only after Release 3, 1994)
DSNUM	Data source number
DSTYPE	Data source type indicates state data organization, hospital association, consortium, and other
HOSPST	State postal code for hospital (e.g., AZ for Arizona)
HOSPSTCO	Modified Federal Information Processing Standards (FIPS) State/county code for hospital, links to Area Resource File
DSNDX	Maximum number of diagnosis codes available from this data source
DSNPR	Maximum number of procedure codes available from this data source
DXSYS	Diagnosis coding system, usually ICD-9-CM
PRSYS	Procedure coding system, usually ICD-9-CM
ADAYWK	Admission day of week
AMONTH	Admission month
PAY1_N	Expected primary payer, nonuniform: (1) Medicare, (2) Medicaid, (3) Blue Cross, (4) commercial, (5) alternative delivery system, incl HMO, (6) self-pay, (7) no charge, (8) Title V, (9) Worker's Compensation, (10) CHAMPUS/CHAMPVA, (11) other government, (12) other
PAY2	Expected secondary payer, uniform: (1) Medicare, (2) Medicaid, (3) private including HMO, (4) self-pay, (5) no charge, (6) other
PAY2_N	Expected secondary payer, nonuniform: (1) Medicare, (2) Medicaid, (3) Blue Cross, (4) commercial, (5) alternative delivery system incl HMO, (6) self-pay, (7) no charge, (8) Title V, (9) Worker's Compensation, (10) CHAMPUS/CHAMPVA, (11) other government, (12) other
Inpatient Stay Core File PR (filename: NIS97_PR.ASC)	
SEQ	Unique sequence number indicates the order in which the data file is sorted (links across files)
PR	Principal and secondary procedures
PRV	Validity flag for principal and secondary procedures – indicates valid, invalid, or missing procedure

Table 4. File Locations of Data Elements in the HCUP NIS (Continued)

Data element	Description (numbers in brackets indicate variable coding)
CD #4	**Compressed file: NIS97CD1.EXE contains:**
Inpatient Stay Core File DX (filename: NIS97_DX.ASC)	
SEQ	Unique sequence number indicates the order in which the data file is sorted (links across files)
DX	Principal and secondary diagnoses
DXV	Validity flag for principal and secondary diagnoses
CD #5	
Two 10% Subsamples (filenames: N97S1ABC.ASC, N97S1_DX.ASC, N97S1_PR. ASC, N97S2ABC.ASC, N97S2_DX.ASC, N97S2_PR. ASC)	All data elements, including discharge weight adjusted for subsampling Data elements for Core Files A, B, C are in one file for each sample; diagnoses and procedures are in separate files
CD #6	
Hospital Weights File (filename: NIS97WT.ASC)	
YEAR	Discharge year (calendar)
HOSPID	HCUP hospital number (links to Inpatient Stay Core File A)
IDNUMBER	AHA ID number (without the leading 6)
AHAID	AHA ID number (with the leading 6 normally included in the AHA identifier), links to AHA Annual Survey
HOSPST	State postal code for hospital
HOSPNAME	Hospital name from AHA Survey
HOSPADDR	Hospital address from AHA Survey
HOSPCITY	Hospital city from AHA Survey
HOSPZIP	Hospital ZIP Code address from AHA Survey
STRAT_ST	Stratum for State-specific weights
STRATUM	Stratum used to post-stratify hospitals
DISCWT_F	Weight to discharges in the frame States
DISCWT_S	Weight to discharges in the State
DISCWT_U	Weight to discharges in the universe
HOSPWT_F	Weight to hospitals in the frame States
HOSPWT_S	Weight to hospitals in the State
HOSPWT_U	Weight to hospitals in the universe
N_DISC_F	Number of frame State discharges in STRATUM
N_DISC_S	Number of State's discharges in STRAT_ST

Table 4. File Locations of Data Elements in the HCUP NIS (Continued)

Data element	Description (numbers in brackets indicate variable coding)
N_DISC_U	Number of universe discharges in STRATUM
N_HOSP_F	Number of frame State hospitals in STRATUM
N_HOSP_S	Number of State's hospitals in STRAT_ST
N_HOSP_U	Number of universe hospitals in STRATUM
S_DISC_S	Number of sample discharges in STRAT_ST
S_DISC_U	Number of sample discharges in STRATUM
S_HOSP_S	Number of sample hospitals in STRAT_ST
S_HOSP_U	Number of sample hospitals in STRATUM
TOTDSCHG	Total hospital discharges
H_BEDSZ	Bedsize of hospital: (1) small, (2) medium, (3) large
H_CONTRL	Control/ownership of hospital: (1) government/nonfederal, (2) private not-for-profit, (3) private investor-owned
H_REGION	Region of hospital: (1) northeast, (2) midwest, (3) south, (4) west
H_LOCTCH	Location/teaching status of hospital: (1) rural, (2) urban nonteaching, (3) urban teaching
H_LOC	Location of hospital: (0) rural, (1) urban
H_TCH	Teaching status of hospital: (0) nonteaching, (1) teaching

Hospital-Level Casemix Counts by DRG, DCCS1, PCCS1
(filenames: N97_HDC1.ASC, N97_HPC1.ASC, N97_HDRG.ASC)

Total number of discharges
Sum of edited length of stay (LOS)
Sum of edited total charges (TOTCHG)
Number of discharges reporting LOS
Number of discharges reporting TOTCHG

NIS Documentation (see Table 5)

NIS Tools (see Table 5)

GETTING STARTED

In order to load NIS data onto your PC, you will need 3,343 megabytes of space available. For individual file requirements, please consult D_Filesp.pdf.

Documentation Using Adobe™ Acrobat™

The documentation for the HCUP Nationwide Inpatient Sample is distributed as portable document format (*.pdf) files along with the Adobe™ Acrobat™ Reader. The documentation on CD #6 contains comprehensive information for Release 6 for the 1997 data.

You must have the Adobe™ Acrobat™ Reader software on your computer to access the NIS documentation. If you don't have Adobe™ Acrobat™ Reader software on your computer, see the README.TXT file on CD #6 for instructions on installing or obtaining the software. The Acrobat™ Reader gives you electronic access to documents in their original form, independent of computer platform or application software. With the Acrobat™ Reader, you can view, navigate through, and print any .pdf file.

The Acrobat™ Reader provided on CD #6 is for IBM-compatible microcomputers running Microsoft Windows. More information and Reader software for other platforms (DOS, Macintosh, Sun Systems, etc.) may be obtained from the Adobe Home Page at http://www.adobe.com/. For further assistance in installing and running the Adobe™ Acrobat™ Reader on your computer platform, please consult your local support personnel.

NIS Documentation and NIS Tools

Table 5 describes the documentation and tools files that can be found on CD #6 and illustrates the structure of the directories and subdirectories on the CD.

Table 5. Directories, Subdirectories and Files on CD #6

\NIS_DOC -- Directory contains all documentation for the NIS, including the Adobe™ Acrobat™ Reader

Documentation files in \NIS_DOC

A_DOCSUM.PDF	Table of Contents for the NIS documentation
B_GENINF.PDF	General Information: An overview of the NIS, file structure, and data elements
C_SUPPMT.PDF	A series of Technical Supplements: TS 1 Data Use Agreement for HCUP Nationwide Inpatient Sample for Release TS 2 Quality Control in HCUP Data Processing TS 3 Mapping Source-Specific Hospital Identifiers to AHA Hospital Identifiers TS 4 Sources of HCUP Data TS 5 Design of the Nationwide Inpatient Sample, Release 1 TS 6 Design of the Nationwide Inpatient Sample, Release 2 TS 7 Design of the Nationwide Inpatient Sample, Release 3 TS 8 Design of the Nationwide Inpatient Sample, Release 4 TS 9 Design of the Nationwide Inpatient Sample, Release 5 TS 10 Design of the Nationwide Inpatient Sample, Release 6 TS 11 Calculating Variances Using Data from the HCUP Nationwide Inpatient Sample TS 12 File Composition for the HCUP Nationwide Inpatient Sample TS 13 HCUP Data Quality Table TS 14 Comparative Analysis of HCUP and NHDS Inpatient Databases
D_FILESP.PDF	File specifications for the 1997 Release 6 NIS files
E_AFILE.PDF	Code Book for the Inpatient Stay Core File A variables
F_BFILE.PDF	Code Book for the Inpatient Stay Core File B variables
G_CFILE.PDF	Code Book for the Inpatient Stay Core File C variables
H_DXFILE.PDF	Code Book for the Inpatient Stay Core File DX variables
I_PRFILE.PDF	Code Book for the Inpatient Stay Core File PR variables
J_ABC1.PDF	Code Book for the Inpatient Stay Sample File 1 variables
K_ABC2.PDF	Code Book for the Inpatient Stay Sample File 2 variables
L_WEIGHT.PDF	Code Book for the Hospital Weights File variables
M_VARNOT.PDF	Notes for all variables (Inpatient Stay and Weights)

\NIS_DOC\READER -- Subdirectory contains the files necessary for installing the Reader; instructions for installing the Adobe™ Acrobat™ Reader appear in a README file under this subdirectory

16bit -- Subdirectory contains the 16 bit version of the Adobe™ Acrobat™ Reader

32bit –Subdirectory contains the 32 bit version of the Adobe™ Acrobat™ Reader

\NISTOOLS -- Directory contains data processing tools that can assist in handling and processing NIS data

\NISTOOLS\SAS -- Subdirectory contains the SAS source code to load the ASCII data files into SAS format, restore the diagnosis and procedure level files to discharge level files, and create a format library

LOAD_A	SAS code to convert the ASCII File A (NIS97_A.ASC) to SAS
LOAD_B	SAS code to convert the ASCII File B (NIS97_B.ASC) to SAS
LOAD_C	SAS code to convert the ASCII File C (NIS97_B.ASC) to SAS
LOAD_DX	SAS code to convert ASCII File DX, diagnoses (NIS97_DX.ASC) to SAS

Table 5. Directories, Subdirectories and Files on CD #6 (Continued)

LOAD_PR	SAS code to convert ASCII File PR, procedures (NIS97_PR.ASC) to SAS
RESTORDX	SAS code to restore File DX (diagnoses) to discharge-level
RESTORPR	SAS code to restore File PR (procedures) to discharge-level
FMTLIB	SAS code to create Inpatient Stay Core File format library
	10% Inpatient Stay Subsample Files
LOADSABC	SAS code to convert the ASCII Subsample Files ABC (NIS97S1_ABC.ASC, NIS97S2_ABC.ASC) to SAS
LOADS_DX	SAS code to convert the ASCII Subsample Files DX, diagnoses (NIS97S1_DX.ASC, NIS97S2_DX.ASC) to SAS
LOADS_PR	SAS code to convert the ASCII Subsample Files PR, procedures (NIS97S1_PR. ASC, NIS97S2_PR.ASC) to SAS
	Hospital Weights File
LOADWT	SAS code to convert Hospital Weight File (NIS97WT.ASC) to SAS
	Hospital-Level Casemix Counts Files
LOADCMDR	SAS code to convert hospital-level casemix counts by DRG10, Diagnosis-Related Groups, Version 10 (N97_HDRG.ASC) to SAS
LOADCMDC	SAS code to convert hospital-level casemix counts by DCCS1, principal diagnosis category (N97_HDC1.ASC) to SAS
LOADCMPC	SAS code to convert hospital-level casemix counts by PCCS1, principal procedure category (N97_HPC1.ASC) to SAS

\NISTOOLS\SPSS -- Subdirectory contains the SPSS source code to load the ASCII data files into SPSS format, restore the diagnosis and procedure level files to discharge level files, and create value labels

	Inpatient Stay Core Files
LOAD_A	SPSS code to convert the ASCII File A (NIS97_A.ASC) to SPSS
LOAD_B	SPSS code to convert the ASCII File B (NIS97_B.ASC) to SPSS
LOAD_C	SPSS code to convert the ASCII File C (NIS97_C.ASC) to SPSS
LOAD_DX	SPSS code to convert the ASCII File DX, diagnoses (NIS97_DX.ASC) to SPSS
LOAD_PR	SPSS code to convert the ASCII File PR, procedures (NIS97_PR.ASC) to SPSS
RESTORDX	SPSS code to restore File DX (diagnoses) to discharge-level
RESTORPR	SPSS code to restore File PR (procedures) to discharge-level
VLABEL	Inpatient Stay Core File value labels
	10% Inpatient Stay Subsample Files
LOADSABC	SPSS code to convert the ASCII Subsample Files ABC (NIS97S1_ABC.ASC, NIS97S2_ABC.ASC) to SPSS
LOADS_DX	SPSS code to convert the ASCII Subsample Files DX, diagnoses (NIS97S1_DX.ASC, NIS97S2_DX.ASC) to SPSS

Table 5. Directories, Subdirectories and Files on CD #6 (Continued)

LOADS_PR	SPSS code to convert the ASCII Subsample Files PR, procedures (NIS97S1_PR.ASC, NIS97S2_PR.ASC) to SPSS
	Hospital Weights File
LOADWT	SPSS code to convert Hospital Weight File (NIS97WT.ASC) to SPSS
	Hospital-Level Casemix Counts Files
LOADCMDR	SPSS code to convert hospital-level casemix counts, by DRG10 , Diagnosis-Related Groups, Version 10 (N97_HDRG.ASC) to SPSS
LOADCMDC	SPSS code to convert hospital-level casemix counts, by DCCS1, principal diagnosis category (N97_HDC1.ASC) to SPSS
LOADCMPC	SPSS code to convert hospital-level casemix counts, by PCCS1, principal procedure category (N97_HPC1.ASC) to SPSS

\NISTOOLS\GENERIC -- Subdirectory contains the record LAYOUTS for all data files and ASCII files containing LABELS for the diagnosis and procedure classification systems

\LABEL -- Subdirectory contains the LABELS for the diagnosis and procedure classification systems

CCS.ASC	Label file for the Clinical Classifications Software (CCSs), a categorization scheme that groups ICD-9-CM diagnosis and procedure codes into mutually exclusive categories (formerly called Clinical Classification for Health Policy Research)
DRG.ASC	Label files for Diagnosis Related Groups (DRGs), multiple versions

\LAYOUT -- Subdirectory contains the record layouts for all data files.

RECL_A	Record layout for Inpatient Stay Core File A (NIS97_A.ASC)
RECL_B	Record layout for Inpatient Stay Core File B (NIS97_B.ASC)
RECL_C	Record layout for Inpatient Stay Core File C (NIS97_C.ASC)
RECL_DX	Record layout for Inpatient Stay Core File DX (NIS97_DX.ASC)
RECL_PR	Record layout for Inpatient Stay Core File PR (NIS97_PR.ASC)
RECLS_ABC	Record layout for Inpatient Stay Subsample ABC Files (N97S1ABC.ASC, N97S2ABC.ASC)
RECLS_DX	Record layout for Inpatient Stay Subsample DX Files (N97S1_DX.ASC, N97S2_DX.ASC)
RECLS_PR	Record layout for Inpatient Stay Subsample PR Files (N97S1_PR.ASC, N97S2_PR.ASC)
RECL_WT	Record layout for Hospital Weights File (NIS97_WT.ASC)
RECLCMDR	Record layout for Hospital-Level Casemix Counts, by DRG10, Diagnosis-Related Groups, Version 10 (N97_HDRG.ASC)
RECLCMDC	Record layout for Hospital-Level Casemix Counts, by DCCS1, principal diagnosis category (N97_HDC1.ASC)
RECLCMPC	Record layout for Hospital-Level Casemix Counts, by PCCS1, principal procedure category (N97_HPC1.ASC)

\CASEMIX -- Directory contains Hospital-Level Casemix Counts files with total number of discharges, sum of edited LOS, sum of edited TOTCHG, number of discharges reporting LOS, number of discharges reporting TOTCHG

Table 5. Directories, Subdirectories and Files on CD #6 (Continued)

N97_HDRG_ASC	by hospital and DRG for 1997
N97_HDC1_ASC	by hospital and principal diagnosis category for 1997
N97_HPC1.ASC	by hospital and principal procedure category for 1997

NIS97WT.ASC -- 1997 Hospital Weights File, includes hospital name, location, linkages to NIS Core files and AHA Survey files, and weights and other variables required to calculate weighted estimates and variances

OTHER HCUP PRODUCTS

The AHRQ Home Page on the World Wide Web is a source of information about HCUP databases and aggregate statistics from HCUP. The address is http://www.ahrq.gov/data/hcup.

Data

NIS Releases are available from the National Technical Information Service (NTIS); call NTIS for prices at 800-553-6847 or 703-605-6000. An online ordering option is available at http://www.ntis.gov/fcpc. All NIS releases are available on CD-ROM; NIS Release 1 may also be purchased on magnetic tape. Order by PB number.

NIS, Release 1, 1988-1992 (PB95-503710)
NIS, Release 2, 1993 (PB96-501325)
NIS, Release 3, 1994 (PB97-500433)
NIS, Release 4, 1995 (PB98-500440)
NIS, Release 5, 1996 (PB99-500480)

The State Inpatient Databases are available directly from selected statewide data sources that contributed data to HCUP as well as from a Central Distributor (for selected states). For more information, send e-mail to hcupsid@medstat.com or visit the HCUP Home Page, http://www.ahrq.gov/data/hcup.

Tools

HCUP Quality Indicators (QIs) are clinical performance measures for use with readily available inpatient data. Methods and software can be downloaded from http://www.ahrq.gov.

Clinical Classifications Software (CCS), formerly known as the Clinical Classifications for Health Policy Research (CCHPRs), are classification systems that group ICD-9-CM diagnoses and procedures into a limited number of clinically meaningful categories. Methods and software can be downloaded from http://www.ahrq.gov/data/hcup/ccs.htm.

Publications

HCUP-3 Research Notes report aggregate statistics and detailed analyses using HCUP data. To request copies, contact the AHRQ Publications Clearinghouse at (800) 358-9295 or send a postcard to: AHRQ Publications Clearinghouse, P.O. Box 8547, Silver Spring, MD 20907 or visit the AHRQ Home Page.

For information on HCUP products:
phone: (301) 594-3075
fax: (301) 594-2166

GENERAL INFORMATION ABOUT HCUP

The Healthcare Cost and Utilization Project, 1988-1997 (HCUP) is a federal-state-industry partnership in health care data. HCUP's objectives are to obtain data from statewide information sources, primarily state governments and hospital associations; design and develop a multistate health care database to be used for health services research and health policy analysis; and release data to a broad set of public and private users.

HCUP has developed two major inpatient databases:

THE STATE INPATIENT DATABASE (SID)

The SID provides uniform data on all inpatient stays in all community hospitals from states with statewide hospital inpatient data systems selected by AHRQ. This database contains common inpatient data elements, variables derived from sensitive data elements, and state-specific variables. Each state database is returned to the respective data source in the uniform HCUP format. Previously, dissemination of SID data was controlled by the state data sources. As of December 1999, 13 SID states make their data available through an AHRQ-sponsored central coordinator and distributor (phone: 805-681-5876, fax: 805-681-5888, e-mail: hcupsid@medstat.com). For more information about the SID, contact:

Agency for Healthcare Research and Quality
2101 East Jefferson Street, Suite 605
Rockville, Maryland 20852
phone: (301) 594-3075
fax: (301) 594-2166
Internet: hcup@ahrq.gov

THE NATIONWIDE INPATIENT SAMPLE (NIS)

The NIS contains inpatient data in a uniform format and is designed to approximate a 20 percent sample of U.S. hospitals. The NIS excludes data elements that could result, directly or indirectly, in the identification of individuals. The NIS is available to researchers who sign a data use agreement restricting the use of the data to research purposes only. The NIS was designed to support the development of national and subnational estimates for costs and utilization of inpatient hospital services.

The NIS will support health policy research on a variety of topics, including:

- the use and cost of hospital services,
- hospital financial distress,
- medical practice variation,
- the effectiveness of medical treatments,
- health care cost inflation,
- the diffusion of medical technology,
- the impact of proposed cost-containment,
- the utilization of health services by special populations, and
- health care reform.

HCUP is directed by the Agency for Healthcare Research and Quality (AHRQ). Development of the HCUP databases was accomplished under contract with The MEDSTAT Group and its subcontractors, Abt Associates and the National Association of Health Data Organizations

(NAHDO). Post-development work, including preparation of the data files for dissemination, was accomplished under contract with Social and Scientific Systems, Inc.

GENERAL INFORMATION ABOUT THE HCUP NATIONWIDE INPATIENT SAMPLE

OVERVIEW

The HCUP Nationwide Inpatient Sample (NIS) is comprised of selected states that have agreed to provide the project with all-payer data on hospital inpatient stays. The NIS contains all discharges from hospitals sampled from these states. Each year of the NIS contains 5-7 million records, and 800 to 1000 hospitals. Table 6 lists the states and years of data in each NIS release. An "X" indicates that data from the state are included.

The NIS contains patient-level clinical and resource-use information included in a typical discharge abstract. The NIS can be linked directly to hospital-level data from the American Hospital Association (AHA) Annual Survey of Hospitals, and to county-level data from the Health Resources and Services Administration, Bureau of Health Professions' Area Resource File (ARF), except for hospitals in Georgia, Hawaii, Kansas, South Carolina, and Tennessee.

The NIS is designed to approximate a 20 percent sample of U.S. community hospitals as defined by the AHA. The AHA defines *community hospitals* as:

> *all nonfederal, short-term, general and other specialty hospitals, excluding hospital units of institutions.*

Included among community hospitals are such specialty hospitals as obstetrics-gynecology, ear-nose-throat, short-term rehabilitation, orthopedic, and pediatric. Not included among community hospitals are long-term hospitals, psychiatric hospitals, and alcoholism and chemical dependency treatment facilities.

This universe of U.S. community hospitals is divided into strata using five hospital characteristics:

- ownership/control,
- bedsize,
- teaching status,
- urban/rural location, and
- U.S. region.

The NIS is a stratified probability sample of hospitals in the frame, with sampling probabilities proportional to the number of U.S. community hospitals in each stratum. The frame is limited by the availability of inpatient data from the data sources. To facilitate the production of national estimates, both hospital and discharge weights are provided, along with information necessary to calculate the variance of estimates.

For a state-by-state description of the NIS, and specific information on the criteria for including hospitals in the NIS, refer to Technical Supplement: *File Composition for the HCUP Nationwide Inpatient Sample*.

Table 6. States in the Frame for the HCUP NIS

State	Calendar Years of Data				
	Release 1		Release 2 and 3	Release 4 and 5	Release 6
	1988	1989-1992	1993- 1994	1995-1996	1997
Arizona	.	X	X	X	X
California	X	X	X	X	X
Colorado	X	X	X	X	X
Connecticut	.	.	X	X	X
Florida	X	X	X	X	X
Georgia					X
Hawaii					X
Illinois	X	X	X	X	X
Iowa	X	X	X	X	X
Kansas	.	.	X	X	X
Maryland	.	.	X	X	X
Massachusetts	X	X	X	X	X
Missouri	.	.	.	X	X
New Jersey	X	X	X	X	X
New York	.	.	X	X	X
Oregon	.	.	X	X	X
Pennsylvania	.	X	X	X	X
South Carolina	.	.	X	X	X
Tennessee	.	.	.	X	X
Utah					X
Washington	X	X	X	X	X
Wisconsin	.	X	X	X	X

STATE-SPECIFIC RESTRICTIONS

Some data sources that contributed data to the NIS imposed restrictions on the release of certain data elements or on the number and types of hospitals that could be included in the database. Table 7 summarizes the state-specific requirements and restrictions for NIS data.

Table 7. Summary of Special Requirements or Restrictions Imposed by Data Sources

State	Requirement/Restriction
AZ	none
CA	none
CO	none
CT	Notify state if more than 50% of hospitals are selected. (Note: fewer than 50% of hospitals were selected for NIS.)
FL	Set admission day of week (ADAYWK) to missing for all records (began in 1993).
GA	Include the HCUP hospital identifier (HOSPID) as the only hospital identifier (e.g., set the AHA identifier, hospital name and address to missing). Set hospital state/county (HOSPSTCO) to missing. Exclude SC hospitals from the sampling frame when there is only 1 SC hospital in a sampling frame stratum (defined by STRATUM). In the Hospital Weights file, set to missing the following hospital characteristics variables when there is only 1 SC hospital in a cell (defined by H_CONTRL, H_BEDSZ, H_LOC, and H_TCH). Variables to be set to missing are: H_CONTRL, H_LOC, H_TCH, H_BEDSZ, H_LOCTCH.
HI	Include the HCUP hospital identifier (HOSPID) as the only hospital identifier (e.g., set the AHA identifier, hospital name and address to missing). Set hospital state/county (HOSPSTCO) to missing. Exclude SC hospitals from the sampling frame when there is only 1 SC hospital in a sampling frame stratum (defined by STRATUM). In the Hospital Weights file, set to missing the following hospital characteristics variables when there is only 1 SC hospital in a cell (defined by H_CONTRL, H_BEDSZ, H_LOC, and H_TCH). Variables to be set to missing are: H_CONTRL, H_LOC, H_TCH, H_BEDSZ, H_LOCTCH.
IL	No more that 40% of IL discharges may appear in any quarter of NIS data. Set all physician identifiers (MDID_S and SURGID_S) to blank. (For 1995, IL did not supply physician identifiers to HCUP.)
IA	none
KS	Include the HCUP hospital identifier (HOSPID) as the only hospital identifier (e.g., set the AHA identifier, hospital name and address to missing). Set hospital state/county (HOSPSTCO) to missing.
MD	none
MA	Physician identifier cannot be released on the NIS. Set all physician identifiers (MDID_S and SURGID_S) to blank.
MO	Exclude MO hospitals from the sampling frame if they did not give their permission to be included in NIS.
NJ	none
NY	none
OR	none

Table 7. Summary of Special Requirements or Restrictions Imposed by Data Sources

State	Requirement/Restriction
PA	Set patient age (AGE and AGEDAY) to the midpoint of five-year intervals for records with the following "sensitive conditions": abortion, AIDS, mental illness, and substance abuse.
SC	Include the HCUP hospital identifier (HOSPID) as the only hospital identifier (e.g., set the AHA identifier, hospital name and address to missing). Set hospital state/county (HOSPSTCO) to missing. Exclude SC hospitals from the sampling frame when there is only 1 SC hospital in a sampling frame stratum (defined by STRATUM). In the Hospital Weights file, set to missing the following hospital characteristics variables when there is only 1 SC hospital in a cell (defined by H_CONTRL, H_BEDSZ, H_LOC, and H_TCH). Variables to be set to missing are: H_CONTRL, H_LOC, H_TCH, H_BEDSZ, H_LOCTCH.
TN	Include the HCUP hospital identifier (HOSPID) as the only hospital identifier (e.g., set the AHA identifier, hospital name and address to missing). Set hospital state/county (HOSPSTCO) to missing. Exclude TN hospitals from the sampling frame when there is only 1 TN hospital in a sampling frame stratum (defined by STRATUM). In the Hospital Weights file, set to missing the following hospital characteristics variables when there is only 1 TN hospital in a cell (defined by H_CONTRL, H_BEDSZ, H_LOC, and H_TCH). Variables to be set to missing are: H_CONTRL, H_LOC, H_TCH, H_BEDSZ, H_LOCTCH.
UT	Physician identifier cannot be released on the NIS. Set all physician identifiers (MDID_S and SURGID_S) to blank.
WA	none
WI	none

FILE STRUCTURE

Using microcomputers to read large data files stored on CD-ROM poses different logistical problems than exist when using mainframe computers to read large data files on magnetic tape or direct access storage devices (DASD). This section briefly describes the evolution of the structure of the HCUP NIS data files from their origin on magnetic tape and then on CD-ROM for microcomputer access.

In general, all releases of the NIS consist of two types of data:

- Inpatient Stay records and
- Hospital Weights.

NIS Inpatient Stay Core Files

Inpatient stay data for the NIS contain four types of ***core data elements***:

- linkage elements,
- patient demographics,

- clinical information, and
- payment information.

Although core data elements were selected in part because of their broad availability across states, not all data elements were available from all states. For example, the number of secondary diagnoses and secondary procedures vary considerably across the states.

Data elements that could directly or indirectly identify individual patients and physicians are excluded. The identities of institutions are available only for data sources that already make that information public or have agreed to its release. (In NIS, Release 1, all data sources agreed to its release. This is not the case for subsequent releases. See Table 7)

In Release 1 there were three NIS Inpatient Stay files:

- ***full NIS files***, which contain 100% of all inpatient records from the sample of NIS hospitals, and
- two non-overlapping ***10% subsamples*** of inpatient records from all NIS hospitals.

In subsequent releases, the inpatient stay files have been restructured to simplify access for microcomputer users but still include:

- ***full NIS files***, which contain 100% of all inpatient records from the sample of NIS hospitals, and
- two non-overlapping ***10% subsamples*** of inpatient records from all NIS hospitals.

Full NIS Files. These NIS files contain 100% of all inpatient records for all community hospitals sampled for the NIS. There was no attempt to delete records from special units (e.g., psychiatric, rehabilitation, long term care) in these community hospitals.

Ten Percent Subsample Files. The 10% Subsamples were drawn by selecting every tenth record from the 100% files beginning with two different starting points randomly selected between 1 and 10. The different starting points guarantee that the Subsamples will not overlap. The two 10% Subsamples can be combined to form a single 20% Subsample of discharges from the NIS.

File Structure for NIS, Release 1 on Magnetic Tapes

The following describes the original structure of the Inpatient Stay Core data files on magnetic tape:

- There is a separate file for each calendar year.
- The unit of observation is the inpatient stay.
- Each calendar-year file contains all discharges for all sampled hospitals for that year.
- Each inpatient stay record contains all core data elements.
- Each inpatient stay record contains SEQ, a unique inpatient stay record identifier.

Figure 1. File Structure for NIS, Release 1, Inpatient Stay Files, 1988-1992, on Magnetic Tape (each cell represents a data file)

1988, all inpatient stay records, all core data elements	1988, 10% Subsample #1, all core data elements
	1998, 10% Subsample #2, all core data elements
1989, all inpatient stay records, all core data elements	1989, 10% Subsample #1, all core data elements
	1989, 10% Subsample #2, all core data elements
1990, all inpatient stay records, all core data elements	1990, 10% Subsample #1, all core data elements
	1990, 10% Subsample #2, all core data elements
1991, all inpatient stay records, all core data elements	1991, 10% Subsample #1, all core data elements
	1991, 10% Subsample #2, all core data elements
1992, all inpatient stay records, all core data elements	1992, 10% Subsample #1, all core data elements
	1992, 10% Subsample #2, all core data elements

File Structure for NIS, Release 1 on CD-ROM

NIS, Release 1, on CD-ROM maintained the tape-based file structure, adapted to the capacity of CD-ROM media. A single CD-ROM can hold about 650 megabytes of data. A complete year of NIS Inpatient Stay data – more than 2 gigabytes in its original form – would not fit onto a single CD-ROM. To accommodate this limitation, each of the calendar-year inpatient stay files of the NIS, Release 1 were split into 4 parts, corresponding to the 4 quarters of the calendar year. (An Inpatient Stay record was allocated to the quarter of its discharge.) Otherwise, the data files remained identical.

With this structure, reading an entire year's worth of NIS data meant reading 4 separate data files located on 4 separate CDs, and production of national estimates still required that some or all of the inpatient stay file be combined with the hospital weights file. To aid the latter task, the hospital weights file was written on the same CD as Quarter 1 of each calendar year.

Figure 2. File Structure for NIS, Release 1, Inpatient Stay Files, 1988-1992, on CD-ROM
(each cell represents a CD)

1988, Quarter 1, all core data elements 1988, Hospital weights file	1988, Quarter 2, all core data elements	1988, Quarter 3, all core data elements	1988, Quarter 4, all core data elements
1989, Quarter 1, all core data elements 1989, Hospital weights file	1989, Quarter 2, all core data elements	1989, Quarter 3, all core data elements	1989, Quarter 4, all core data elements
1990, Quarter 1, all core data elements 1990, Hospital weights file	1990, Quarter 2, all core data elements	1990, Quarter 3, all core data elements	1990, Quarter 4, all core data elements
1991, Quarter 1, all core data elements 1991, Hospital weights file	1991, Quarter 2, all core data elements	1991, Quarter 3, all core data elements	1991, Quarter 4, all core data elements
1992, Quarter 1, all core data elements 1992, Hospital weights file	1992, Quarter 2, all core data elements	1992, Quarter 3, all core data elements	1992, Quarter 4, all core data elements
1988, 10% Subsample #1, all core data elements 1988, 10% Subsample #2, all core data elements 1988, Hospital Weights file			
1989, 10% Subsample #1, all core data elements 1989, 10% Subsample #2, all core data elements 1989, Hospital Weights file			
1990, 10% Subsample #1, all core data elements 1990, 10% Subsample #2, all core data elements 1990, Hospital Weights file			
1991, 10% Subsample #1, all core data elements 1991, 10% Subsample #2, all core data elements 1991, Hospital Weights file			
1992, 10% Subsample #1, all core data elements 1992, 10% Subsample #2, all core data elements 1992, Hospital Weights file			
NIS Documentation NIS Tools			

File Structure for NIS, Releases 2-6 on CD-ROM – *Quick Reference* Format

As described above, our first foray into processing data on CD-ROM using microcomputers was limited by our experience with mainframe computers and magnetic tape access. Experience in using NIS data on CD-ROM led us to adopt an alternative file structure for NIS, Release 2. This file structure has been maintained for all subsequent NIS releases. NIS, Release 2 is only available on CD-ROM. The magnetic tape product was eliminated because of the limited demand for magnetic tape products for Release 1.

The goal of the new file structure is to simplify data access – to make it possible to do more from a single CD. This new structure, referred to as the *Quick Reference* (QR) format, is illustrated below:

Figure 3. File Structure for NIS, Releases 2-6, Inpatient Stay Files, 1993-1997, on CD-ROM
(each cell represents a CD)

CD #1 Inpatient Stay Core File A (most commonly used core data elements) Hospital Weights File
CD #2 Inpatient Stay Core File B (less commonly used core data elements)
CD #3 Inpatient Stay Core File C (less commonly used data elements) Inpatient Stay Core File PR (normalized procedure file)
CD #4 Inpatient Stay Core File DX (normalized diagnosis file)
CD #5 10% Subsample #1, data elements from Files A, B, and C, combined 10% Subsample #1 Core File PR (normalized procedure file) 10% Subsample #1 Core File DX (normalized diagnosis file) 10% Subsample #2, data elements from Files A, B, and C, combined 10% Subsample #2 Core File PR (normalized procedure file) 10% Subsample #2 Core File DX (normalized diagnosis file)
CD #6 Hospital-level Casemix Counts Hospital Weights file NIS Documentation NIS Tools

FILE LOCATIONS OF DATA ELEMENTS IN THE NIS

Full NIS Files

NIS, Release 1, has one Inpatient Stay Core File per year that contains all core data elements.

In subsequent releases of the full NIS, core data elements are distributed across Inpatient Stay Core Files A, B, and C roughly according to their usefulness or expected frequency of use. Core File A contains the most commonly used core data elements (e.g., principal diagnosis, principal procedure, length of stay, total charges, age, sex, etc.) and the discharge weight (DISCWT_U) most commonly used to produce national estimates. Core Files B and C contain less frequently used elements. Table 8 shows the distribution of data elements across the QR files. The full complement of diagnoses and procedures are stored in normalized files, described below.

Table 8. File Locations of Data Elements in the HCUP NIS

Data Element	Description	Inpatient Stay Core Files					
		NIS-R1 1988-1992	NIS-R2, NIS-R3, NIS-R4, NIS-R5, NIS-R6 1993, 1994, 1995, 1996, 1997				
			A	B	C	DX	PR
Linkage Elements							
HOSPID	HCUP hospital number	X	X	.	.	.	.
PROCESS	Processing number assigned for tracking records throughout data processing	X	.	.	X	.	.
SEQ	Unique sequence number indicates the order in which the data file is sorted	X	X	X	X	X	X
SEQ_SID	Unique sequence number assigned for processing original SID data file (Release 3, 1994 only)	.	.	.	X	.	.
YEAR	Calendar year of discharge	X	.	.	.	.	.
Physician Identifiers							
MDID_S	Synthetic attending physician number	X	.	X	.	.	.
SURGID_S	Synthetic primary surgeon number	X	.	X	.	.	.
Data Source Identifiers							
DSHOSPID	Hospital number as received from the data source	X	.	X	.	.	.

Table 8. File Locations of Data Elements in the HCUP NIS

Data Element	Description	Inpatient Stay Core Files					
		NIS-R1 1988-1992	NIS-R2, NIS-R3, NIS-R4, NIS-R5, NIS-R6 1993, 1994, 1995, 1996, 1997				
			A	B	C	DX	PR
DSNUM	Data source number	X	.	.	X	.	.
DSTYPE	Data source type indicates state data organization, hospital association, consortium, and other	X	.	.	X	.	.
Area Identifiers							
HOSPST	State postal code for hospital	X	.	.	X	.	.
HOSPSTCO	Modified FIPS state/county code for hospital	X	.	.	X	.	.
ZIPINC4	Median income for patient's zip code (in 4 categories)	X	X	.	.	.	.
ZIPINC8	Median income for patient's zip code (in 8 categories)	X	.	X	.	.	.
Patient Demographics							
AGE	Age in years at admission	X	X	.	.	.	.
AGEDAY	Age in days (coded only when the age in years is less than 1)	X	X	.	.	.	.
RACE	Race includes white, black, Hispanic, Asian, Pacific Islander, and Native American	X	X	.	.	.	.
SEX	Sex	X	X	.	.	.	.
Clinical Information							
DX1	Principal diagnosis	X	X	.	.	.	.
DX2-DX15	Secondary diagnoses	X	.	.	.	.	.
DX	Principal and secondary diagnoses	.	.	.	.	X	.
DXV1	Validity flag for principal diagnosis – indicates valid, invalid, or missing diagnosis	X	X	.	.	.	.

Table 8. File Locations of Data Elements in the HCUP NIS

Data Element	Description	Inpatient Stay Core Files					
		NIS-R1 1988-1992	NIS-R2, NIS-R3, NIS-R4, NIS-R5, NIS-R6 1993, 1994, 1995, 1996, 1997				
			A	B	C	DX	PR
DXV2-DXV15	Validity flag for secondary diagnoses – indicates valid, invalid, or missing diagnosis	X	.	.	.	.	.
DXV	Validity flag for principal and secondary diagnoses	.	.	.	.	X	.
NDX	Number of diagnoses coded on the original record	X	X	.	.	.	.
DSNDX	Maximum number of diagnosis codes available from this data source	X	.	.	X	.	.
PR1	Principal procedures	X	X	.	.	.	.
PR2-PR15	Secondary procedures	X	.	.	.	.	.
PR	Principal and secondary procedures	.	.	.	.	.	X
PRV1	Validity flag for principal procedure – indicates valid, invalid, or missing procedure	X	X	.	.	.	.
PRV2-PRV15	Validity flag for secondary procedures – indicates valid, invalid, or missing procedure	X	.	.	.	.	.
PRV	Validity flag for principal and secondary procedures – indicates valid, invalid, or missing procedure	.	.	.	.		X
NPR	Number of procedures coded on the original record	X	X	.	.	.	.
DSNPR	Maximum number of procedure codes available from this data source	X	.	.	X	.	.
DXSYS	Diagnosis coding system, usually ICD-9-CM	X	.	.	X	.	.
PRSYS	Procedure coding system, usually ICD-9-CM	X	.	.	X	.	.

Table 8. File Locations of Data Elements in the HCUP NIS

Data Element	Description	Inpatient Stay Core Files					
		NIS-R1 1988-1992	NIS-R2, NIS-R3, NIS-R4, NIS-R5, NIS-R6 1993, 1994, 1995, 1996, 1997				
			A	B	C	DX	PR
DRG	DRG in use on discharge date	X	X	.	.	.	.
MDC	MDC in use on discharge date	X	X	.	.	.	.
DRGVER	Grouper version in use on discharge date	X	X	.	.	.	.
DRG10	DRG Version 10 (October 1992)	X	X	.	.	.	.
MDC10	MDC Version 10 (October 1992)	X	X	.	.	.	.
DCCHPR1	Clinical Classification for Health Policy Research (CCHPR) for principal diagnosis	X	X	.	.	.	.
PCCHPR1	CCHPR for principal procedure	X	X	.	.	.	.
NEOMAT	Neonatal/maternal flag	X	.	X	.	.	.
Days and Dates							
ADAYWK	Admission day of week	X	.	.	X	.	.
AMONTH	Admission month	X	.	.	X	.	.
DQTR	Discharge quarter	X	X	.	.	.	.
LOS	Length of stay, edited	X	X	.	.	.	.
LOS_X	Length of stay, unedited	X	.	X	.	.	.
PRDAY1	Days from admission of principal procedure	X	.	X	.	.	.
Admission/Discharge Status							
ASOURCE	Admission source	X	.	X	.	.	.
ATYPE	Admission type	X	.	X	.	.	.
DIED	Indicates in-hospital death	X	X	.	.	.	.

Table 8. File Locations of Data Elements in the HCUP NIS

Data Element	Description	Inpatient Stay Core Files					
		NIS-R1 1988-1992	NIS-R2, NIS-R3, NIS-R4, NIS-R5, NIS-R6 1993, 1994, 1995, 1996, 1997				
			A	B	C	DX	PR
DISP	Disposition of patient (discharge status)	X	X	.	.	.	.
Payment Information							
PAY1	Expected primary payer	X	X	.	.	.	.
PAY1_N	Expected primary payer (more detailed than PAY1)	X	.	.	X	.	.
PAY2	Expected secondary payer	X	.	.	X	.	.
PAY2_N	Expected secondary payer (more detailed than PAY2)	X	.	.	X	.	.
TOTCHG	Total charges, edited	X	X	.	.	.	.
TOTCHG_X	Total charges, unedited	X	.	X	.	.	.
Weights							
DISCWT_U	Weight to discharges in universe	.	X	.	.	.	.

Hospital-level Casemix Counts

These are unweighted hospital-level summary files generated from the full NIS. These files were generated to provide a readily available source of frequently used aggregates by hospital and to eliminate the need for analysts to generate this information repeatedly.

Table 9 shows the types of casemix counts generated for the NIS.

Table 9. Hospital-level Casemix Counts

Type of Casemix Counts	Levels of Aggregation
Total number of discharges	By hospital and DRG10
Sum of edited LOS	By hospital and DCCHPR1 (principal diagnosis)
Sum of edited TOTCHG	By hospital and PCCHPR1 (principal procedure)
Number of discharges reporting LOS	
Number of discharges reporting TOTCHG	

Ten Percent Subsample Files

Inpatient Stay Files A, B, and C were combined for the two 10% subsamples, which also include the discharge weight most commonly used to produce national estimates (multiplied by 10, for the subsamples). The full complement of diagnoses and procedures are stored in normalized files, described below.

Normalized Diagnoses (Core File DX) and Procedures (Core File PR)

The greatest economies in terms of data storage came from restructuring the diagnosis and procedure vectors. In the original HCUP structure, each inpatient stay record carried space for 15 diagnoses, 15 procedures, and their associated validity flags. However, 90% of inpatient stay records have 7 or fewer diagnoses and 3 or fewer procedures recorded, so the original structure was heavy with wasted space.

Restructuring the diagnosis and procedure portions of the NIS inpatient stay records dramatically reduced the storage space required. In addition, the new structure promotes speedier processing of diagnosis and procedure arrays. It simply takes less time to read the new *skinny* files.

Figure 4 illustrates the format of diagnoses in the original HCUP NIS structure, which contains 15 diagnoses. (In reality, each state reports different numbers of diagnoses and the NIS accommodates up to 15). SEQ is the unique inpatient record identifier; NDX is a count of the number of diagnoses on a record.

Figure 4. Diagnoses in the Original HCUP format

SEQ	NDX	DX1-DX15														DXV1-DXV15												
1	4	A	B	C	D											1	0	0	0									
2	2	E	F													0	1											
3	3		C	E													0	0											

Figure 5 illustrates the restructured *normalized* File DX. Inpatient Stay Core File DX contains only three variables: SEQ, DX, and DXV. For each inpatient stay record, uniquely identified by SEQ, File DX includes one record for the principal diagnosis and as many additional records as there are secondary diagnoses recorded. Thus, an inpatient stay record with 1 diagnosis (NDX=1) has been converted to 1 SEQ-DX record; an inpatient stay record with 2 diagnoses (NDX=2) has been converted to 2 SEQ-DX records, and so forth.

Figure 5. Normalized File DX

SEQ	DX	DXV
1	A	1
1	B	0
1	C	0
1	D	0
2	E	0
2	F	1
3		
3	C	0
3	E	0

The diagnoses in File DX maintain their relative positions from the original diagnosis array. This is a critical feature for preserving the identity of the principal diagnosis. In File DX, the first diagnosis record for the SEQ is ***always*** the principal diagnosis (DX1). When the principal diagnosis is missing, which occurs rarely, the first record for that SEQ in File DX preserves the blank value for data element DX.

File PR was created using the same logic. The restructured normalized File PR contains only three variables: SEQ, PR, and PRV, and it includes as many SEQ-PR records as the number of procedures (NPR) in the original HCUP database. Again, the procedures maintain their relative positions from the original procedure array. Unlike the principal diagnosis, the principal procedure is often missing (in more than 30% of stays) because many patients do not undergo a procedure in the hospital. File PR preserves the blank principal procedure.

NIS Hospital Weights Files

The NIS Hospital Weights file contains one observation for each hospital included in the NIS. The unit of observation is the *hospital*. The HCUP hospital identifier (HOSPID) provides the linkage between the NIS Inpatient Stay file and the Hospital Weights file for each calendar year.

In combination, the Inpatient Stay file and the Hospital Weights file contain all the data elements required to produce national estimates, including the variance of estimates. However, production of national estimates requires that some or all of the inpatient stay file be combined with the hospital weights file.

Discharge weights can be multiplied by 10 for use with the Subsample files. Hospital weights do not require adjustment because every NIS hospital appears in each 10% Subsample. A discharge weight variable (D10CWT_U), appropriate for use with the subsamples, is included in

the Subsample files. For detailed information about use of discharge and hospital weights, see the Technical Supplements on *Design of the HCUP Nationwide Inpatient Sample*.

The Hospital Weights files contain five types of data elements:

- linkage elements,
- strata definitions,
- sample weights,
- key hospital characteristics, and
- variables required to calculate variance estimates.

Table 10 lists the data elements in the Hospital Weights File.

The linkage elements include the hospital identifier assigned by the AHA, providing a direct link to the AHA Annual Survey of Hospitals, which can be obtained directly from the American Hospital Association.

Table 10. Data Elements for Hospital Weights File in the HCUP NIS

Data Element	Description
Linkage Elements	
YEAR	Discharge year (calendar)
HOSPID	HCUP hospital number
IDNUMBER	AHA ID number (without the leading 6)
AHAID	AHA ID number (with the leading 6 normally included in the AHA identifier)
HOSPST	State postal code for hospital
HOSPNAME	Hospital name from AHA Survey†
HOSPADDR	Hospital address from AHA Survey†
HOSPCITY	Hospital city from AHA Survey†
HOSPZIP	Hospital zip code address from AHA Survey†
Strata Definitions	
STRAT_ST	Stratum for state-specific weights
STRATUM	Stratum used to post-stratify hospitals
Sample Weights	
DISCWT_F	Weight to discharges in the frame states
DISCWT_S	Weight to discharges in the state
DISCWT_U	Weight to discharges in the universe
D10CWT_U	Weight to discharges in the universe) use with 10% Subsamples†
HOSPWT_F	Weight to hospitals in the frame states
HOSPWT_S	Weight to hospitals in the state

Table 10. Data Elements for Hospital Weights File in the HCUP NIS

Data Element	Description
HOSPWT_U	Weight to hospitals in the universe
Key Hospital Characteristics	
H_BEDSZ	Bedsize of hospital‡
H_CONTRL	Control/ownership of hospital‡
H_REGION	Region of hospital‡
H_LOCTCH	Location/teaching status of hospital‡
H_LOC	Location (urban/rural) of hospital‡
H_TCH	Teaching status of hospital‡

Table 10. Data Elements for Hospital Weights File in the HCUP NIS

Data Element	Description
Data Elements Required to Calculate Variances	
N_DISC_F	Number of frame state discharge in STRATUM
N_DISC_S	Number of state's discharges in STRAT_ST
N_DISC_U	Number of universe discharge in STRATUM
N_HOSP_F	Number of frame state hospitals in STRATUM
N_HOSP_S	Number of state's hospitals in STRAT_ST
N_HOSP_U	Number of universe hospitals in STRATUM
S_DISC_S	Number of sample discharges in STRAT_ST
S_DISC_U	Number of sample discharges in STRATUM
S_HOSP_S	Number of sample hospitals in STRAT_ST
S_HOSP_U	Number of sample hospitals in STRATUM
TOTDSCHG	Total number of discharges in the hospital

† These data elements were not included in Release 1.
‡ These data elements had different names in Release 1.

Advantages of the *Quick Reference* Format

There are several advantages to users of data in the *Quick Reference* format.

- First, many common applications for the full NIS will require only File A, which can be read from a single CD.
- Second, many efforts to develop national estimates can be accomplished using a single data file (File A for the full NIS or either 10% Subsample file) that already contains the discharge weight pre-linked from the hospital weights file.
- Third, applications that require subsetting inpatient stays based on diagnostic or procedure criteria will be more efficient when using the normalized File DX or File PR to identify the subset, which can then be extracted from Files A, B, and/or C using the SEQ linkage.

NIS Documentation and NIS Tools for the *Quick Reference* Format

The codebooks, which contain coding and summary statistics for each data element, were generated from the full NIS inpatient stay-level files in original format. The result is that we provide coding and statistics for each diagnosis – DX1, DX2, DX3, etc. – and for each procedure – PR1, PR2, PR3, etc. – not for the normalized DX and PR vectors. Summary statistics for the subsample files and for the normalized diagnosis and procedure files are not available.

NIS Tools include source code for SAS and SPSS users to reassemble diagnoses and procedures into their original formats. Diagnosis or procedure arrays reassembled in this manner can then be combined as needed with Files A, B, or C using the linkage by SEQ.

HCUP CODING

The following objectives guided the definition of variables included in the HCUP NIS:

- Make the database as usable as possible without extensive editing by analysts.
- Retain the largest amount of information available from the original sources, while still maintaining consistency among sources.
- Structure the information for efficient storage, manipulation, and analysis.
- Set variable attributes (type and length) to accommodate all expected discharge data. The required characteristics were determined from:
    - The actual characteristics of state and hospital association data tabulated in the HCUP Feasibility Study (*AHRQ Hospital Cost Database Feasibility Study,* Contract No. 282-90-0029).
    - National standards, including the Uniform Hospital Discharge Data set (UHDDS), Uniform Bill 1982 (UB-82), and Uniform Bill 1992 (UB-92).

Coding of Variables

Variables are coded as shown in Table 11:

Table 11. Coding Conventions

Values have been:	Examples of variables:
Retained in the form provided by the data source	Diagnosis and procedure codes
Encrypted into synthetic values	Physician identifiers
Recoded into uniform coding schemes	Sex, race, expected primary pay source
Calculated (when possible)	Age, length of stay, day of principal procedure
Assigned using external algorithms	Diagnosis Related Groups (DRGs), Clinical Classification for Health Policy Research (CCHPR)

Attributes of Variables

Variables are defined as numeric or character.

- Numeric format is used for data elements that are reasonable to express numerically (e.g., age of patient); and for most categorical variables (e.g., sex of patient).

  Categorical variables are expressed in numeric format, because that format:
    - facilitates logical comparisons of indicator variables and
    - permits flexibility in the creation of summary statistics.

- Character format is used for data elements that contain alphanumeric characters not amenable to recoding. Some variables are expressed in character format because:
  - the alphanumeric data have a recognized significance that must be preserved (e.g., ICD-9-CM diagnosis and procedure codes); and
  - there is no reasonable conversion to numeric coding (e.g., encrypted physician identifiers).

- To save storage space, variable lengths are limited to what is necessary to accommodate the expected data.

Missing Values

Special missing values have been used in HCUP NIS variables to indicate details of data availability and quality. Missing values differ depending on whether you have obtained NIS data in SAS transport, EBCDIC or ASCII formats. NIS, Release 1 data (1988-1992) are available in all three formats. SAS transport and EBCDIC formats are used for data on magnetic tapes. ASCII format is used for data on CD-ROM. Beginning with NIS, Release 2 (1993), NIS data are available only in ASCII format on CD-ROM.

- **Missing Data**

  When:
  - the source has defined an explicit value as unknown or unavailable
  - the source uses a default missing value to indicate missing data
  - exploratory statistics show an undocumented value with a frequency suggestive of a missing value, *and*

  it is a commonly used missing value (e.g., blank, zero, or 9-filled), or when contacted, the source confirms that the value is unknown or unavailable

  the following missing values are assigned:

  SAS
  - a value of "." for numeric variables
  - " " (blank) for character variables

  EBCDIC/ASCII
  - a negative 9-filled value (-9, -99, -999, etc.) for numeric variables
  - " " (blank) for character variables

- **Invalid Data**

  When the source data contain undocumented, out-of-range, or invalid values, e.g., a negative value for age, or an alpha character in a numeric field, the following missing values are assigned:

  SAS
  - ".A" for numeric variables

  EBCDIC/ASCII
  - a negative 8-filled value (-8, -88, etc.) for numeric variables

  For diagnoses and procedures, an invalid code is retained in the diagnosis/procedure variable. The presence of the invalid code is recorded in the indicator variable (DXVnn/PRVnn) associated with that diagnosis or procedure.

- **Data Unavailable from Source**

  When a source does not provide a data element, the following missing values are assigned:

  SAS
  - ".B" for numeric variables

  EBCDIC/ASCII
  - a negative 7-filled value (-7, -77, etc.) for numeric variables

- **Inconsistent Data**

  Related variables within the same record were checked for logical consistency, e.g., a procedure of *hysterectomy* reported with a sex of *male* is inconsistent. When such inconsistencies were identified, the following missing values were assigned:

  SAS
  - ".C" for numeric variables

  EBCDIC/ASCII
  - a negative 6-filled (-6, -66, etc.) value for numeric variables

  For diagnoses and procedures, an inconsistent code is retained in the diagnosis/procedure variable. The presence of the inconsistent code is recorded in the indicator variable (DXVnn/PRVnn) associated with that diagnosis or procedure.

See the *Quality Control* section below for details.

QUALITY CONTROL

This section describes the procedures used to assess data quality for each data source participating in HCUP.

Quality Control Philosophy

Edit procedures were applied to the HCUP NIS. Editing followed explicit rules:

- Make the data usable without extensive further editing.
- Confirm that data values are valid, internally consistent, and consistent with established norms, when feasible.
- Use some edit procedures to set questionable and inconsistent values to inconsistent (.C or negative 6-filled). Use other edit procedures only to tabulate edit failures. Use the latter to evaluate whether systematic problems exist.
- Never "fix" or impute data. Set invalid or inconsistent values to missing or, for diagnoses and procedures, set flags to indicate invalid or inconsistent codes. This preserves the analyst's ability to investigate data anomalies.
- Some data elements are more important than others because:

  - they are coded more reliably because they relate to reimbursement; and
  - without these variables, a discharge record is not useful for most analytic purposes.

  Therefore, values of these data elements should be retained even in the presence of conflicting information. In order of importance, these data elements are:

  1. Discharge date (and within discharge date: year, month, and day)
  2. Admission date
  3. Principal diagnosis

  Note: Although discharge date and admission date are not part of the NIS, these variables were employed to compute length of stay.

- Tabulate instances of edit failures and use these to assess data quality for each data source.

Quality Review

The following statistics were reviewed by an independent contractor for each year and data source (or for each different layout if the source changed file layouts during the year):

- For all numeric variables – means, number of missing and nonmissing values, minimum, and maximum.
- For all categorical and some continuous variables – frequency distributions.
- For closely related variables (e.g., age in years compared to age in days) – cross-frequencies.

In addition, the following statistics were reviewed for each hospital:

- The number of discharges in consecutive quarters – to identify large fluctuations unexplained by closures or mergers of hospitals.

- The distributions of edited length of stay, total charges, and charge per day – to identify atypical distributions.

For details, see Technical Supplement: *Quality Control in HCUP Data Processing*.

Automated Quality Control Procedures

The following procedures were applied to each inpatient discharge record:

- *To assess validity of values* –

  For numeric data:
  - Verify numeric data as numeric.
  - Check the range against legal values documented by the data source.
  - Check the range against standard norms (e.g., length of stay is a non-negative value; age in years is between 0 and 124, the maximum allowed by the DRG grouper).
  - Check the values against the maximum allowed for the variable (e.g., length of stay less than 32,767).

  For character data:
  - Verify against norms, when feasible (e.g., diagnosis codes, procedure codes, patient zip codes).

- *To assess internal consistency* –

  Compare values of related variables (e.g., a procedure of *hysterectomy* should appear with a sex of *female;* admission date should occur *before* discharge date).

  If an inconsistency involves a critical variable (such as discharge date, admission date, or principal diagnosis), retain the critical variable according to the established hierarchy. For example:

  - If discharge date falls before admission date, retain discharge date and set admission date and length of stay to inconsistent (negative 6-filled or .C).

  - If discharge date is invalid (e.g., February 30), retain discharge quarter and discharge year.

- *To assess consistency with established norms* –
  Compare values to an established norm (e.g., maternal diagnoses should occur with an age between 10 and 55 years).

Data Quality Variables

Three types of data quality variables were created during inpatient discharge data processing.

Diagnosis and Procedure Validity Flags – DXVnn and PRVnn. These flags indicate invalid or inconsistent data in the associated diagnosis and procedure variables. Original values of the diagnoses and procedures are maintained. DXVnn and PRVnn have the following values:

0	Diagnosis or procedure code is valid and consistent.
1	Diagnosis or procedure code is invalid as of the discharge date, plus or minus three months (to allow for anticipation of or lags in response to official ICD-9-CM coding changes).
.C or -6	Diagnosis or procedure code is inconsistent with age or sex on the same record.
. or -9	No diagnosis or procedure coded.

Neonatal/Maternal Indicator Flag – NEOMAT. This variable identifies discharges with neonatal and/or maternal diagnoses or procedures. Maternal diagnoses and procedures do not necessarily result in a delivery. For a definition of neonatal and maternal diagnoses and procedures, refer to Technical Supplement: *Quality Control in HCUP Data Processing*.

NEOMAT has the following values:

0	No neonatal/maternal diagnoses or procedures.
1	Maternal diagnoses and/or procedures are present.
2	Neonatal diagnoses are present.
3	Both neonatal and maternal diagnoses and/or procedures are present.

SELECTION OF HOSPITALS INCLUDED IN THE NIS

The hospital universe is defined by all hospitals that were open during any part of each calendar year and were designated as community hospitals in the AHA Annual Survey of Hospitals.

For more information on how hospitals in the data were mapped to hospitals as defined by the AHA, refer to Technical Supplement: *Mapping Source-Specific Hospital Identifiers to AHA Hospital Identifiers*.

For a list of all data sources, refer to Technical Supplement: *Sources of HCUP Data*.

For more detailed descriptions of the sampling design, refer to Technical Supplements on *Design of the HCUP Nationwide Inpatient Sample.*

Stratification Variables

To help ensure generalizability, five hospital sampling strata were defined based on hospital characteristics contained in the AHA Annual Survey of Hospitals. The stratification variables were:

- *Geographic Region* – Northeast, Midwest, West, and South. This is an important stratifier because practice patterns have been shown to vary substantially by region. For example, lengths of stay tend to be longer in East Coast hospitals than in West Coast hospitals.

- *Ownership* – Government nonfederal, private not-for-profit, private investor-owned. These types of hospitals tend to have different missions and different responses to government regulations and policies.

- *Location* – Urban or rural. Government payment policies often differ according to this designation. Also, rural hospitals are generally smaller and offer fewer services than urban hospitals.

- *Teaching Status* – Teaching or nonteaching. A hospital is considered to be a teaching hospital if it has either 1) an AMA-approved residency program, or 2) membership in the Council of Teaching Hospitals (COTH). Teaching hospitals have a different mission than nonteaching hospitals. Additionally, financial considerations differ between teaching and nonteaching hospitals. Currently, the Medicare DRG payments are uniformly higher to teaching hospitals than to nonteaching hospitals.

- *Bedsize* – Small, medium, and large. Bedsize categories are specific to the hospital's location and teaching status as shown in Table 12.

Table 12. Bedsize Categories

Location and Teaching Status	Bedsize		
	Small	Medium	Large
Rural	1-49	50-99	100+
Urban, nonteaching	1-99	100-199	200+
Urban, teaching	1-299	300-499	500+

Rural hospitals were not split according to teaching status because rural teaching hospitals were rare. The cut-off points for the bedsize categories are consistent with those used in *Hospital Statistics,* published annually by the AHA.

Hospital Sampling Frame

For each year, the *universe* of hospitals was established as all community hospitals located in the U.S. The NIS *sampling frame* was constructed from the subset of universe hospitals that released their all-payer inpatient data for government use. At the time the sample was drawn, AHRQ had agreements with 22 data sources that maintained statewide, all-payer discharge data in 1997. However, the first release of the NIS included only 8 states in 1988 and 11 states in 1989-1992. An additional 6 states were included in the second and third releases of the NIS for 1993 and

1994, respectively. Two more states were included in the fourth and fifth releases of the NIS for 1995 and 1996.

The following additional restrictions were placed on the sampling frame:

- Certain states stipulated limitations on the percentage of discharges that could be included in the NIS database for any calendar year. For details, see the Technical Supplements on *Design of the HCUP Nationwide Inpatient Sample* and *File Composition for the HCUP Nationwide Inpatient Sample*.

- Only community hospitals provided by the data sources could be included in the sampling frame. If an AHA community hospital was not provided by a participating state, that hospital was eliminated from the sampling frame (but not from the universe).

Hospital Sampling Procedure

Once the universe of hospitals was stratified, up to 20 percent of the total number of U.S. hospitals was randomly selected within each stratum. If too few frame hospitals were in the stratum, then all frame hospitals were selected for the NIS. Frame hospitals within a given stratum all had an equal chance of being included in the sample.

The 1993 sample was drawn by a procedure that retained most of the 1992 hospitals, while allowing hospitals new to the frame an opportunity to enter the 1993 NIS. In particular, hospitals in six states entered the 1993 frame that were not in the 1992 frame (CT, KS, MD, NY, OR, and SC).

The 1994 sample was drawn by a procedure that retained most of the 1993 hospitals, while allowing hospitals new to the frame an opportunity to enter the 1994 NIS. No additional states were added to the 1994 frame, but even in frame states that were present in the 1993 sample, hospitals that opened in 1994 needed a chance to enter the sample. Also, hospitals that changed strata between 1993 and 1994 were considered new to the 1994 frame.

The 1995 sample was drawn by a procedure that retained most of the 1994 hospitals, while allowing hospitals new to the frame an opportunity to enter the 1995 NIS. In particular, hospitals in two states entered the 1995 frame that were not in the 1994 frame (MO and TN). Even in frame states that were present in the 1994 sample, hospitals that opened in 1995 needed a chance to enter the sample. Also, hospitals that changed strata between 1994 and 1995 were considered new to the 1995 frame. The same procedure was followed for the 1996 and 1997 frames.

Zero-Weight Hospitals

The 1993-1997 samples contain no zero-weight hospitals. For a description of zero-weight hospitals in the 1988-1992 sample, see Technical Supplement: *Design of the HCUP Nationwide Inpatient Sample, Release 1*.

Final Hospital Sample

Table 13 describes the number of hospitals in the hospital universe; in the sampling frame; and in the NIS sample, by year. In this table, the NIS sample excludes the zero-weight hospitals which were kept in the database only to maintain a longitudinal analysis file, thus the numbers for hospital sample and number of discharges differ from Table 3.

Table 13. NIS Universe, Sampling Frame, and Sample

Calendar Year	Hospital Universe	Sampling Frame	NIS Sample	
			Hospital Sample	Number of Discharges
NIS, Release 1				
1988	5,607	1,247	758	5,242,904
1989	5,548	1,658	875	6,067,667
1990	5,468	1,620	861	6,156,638
1991	5,412	1,604	847	5,984,270
1992	5,334	1,591	838	6,008,001
NIS, Release 2				
1993	5,313	2,168	913	6,538,976
NIS, Release 3				
1994	5,290	2,135	904	6,385,011
NIS, Release 4				
1995	5,260	2,284	938	6,688,692
NIS, Release 5				
1996	5,182	2,268	906	6,542,069
NIS, Release 6				
1997	5,113	2,452	1,012	7,148,420

SAMPLING WEIGHTS

Although the sampling design was simple and straightforward, it is necessary to incorporate sample weights to obtain state and national estimates. Sample weights were developed separately for hospital- and discharge-level analyses. Hospital-level weights were developed to weight NIS sample hospitals to the state, frame, and universe. Likewise, discharge-level weights were developed to weight NIS sample discharges to the state, frame, and universe.

The hospital weights are useful for producing hospital-level statistics for analyses that use the *hospital* as the unit of analysis. The discharge weights are useful for producing discharge-level statistics for analyses that use the *discharge* as the unit of analysis. These would be used to weight the sample data in estimating population statistics.

For more information on sampling weights and the calculation of the variance of estimates, see the Technical Supplements on *Design of the HCUP Nationwide Inpatient Sample.*